Chapter 7

INTRODUCTION TO PROBABILITY MODELS

Most real–life problems contain elements of uncertainty. In some models we may introduce random elements to account for uncertainties in human behavior. In other models we may be unsure of the exact physical parameters of a system, or we may be unsure of the exact physical laws that govern its dynamics. It has even been suggested in some cases that physical parameters and physical laws are essentially random; for example, in quantum mechanics. Sometimes probabilities are introduced into a model as a matter of convenience, sometimes as a matter of necessity. In either case, it is here in the realm of probability that mathematical modeling becomes most interesting and useful.

Probability is a familiar and intuitive idea. In this chapter we begin our treatment of probability models. We do not assume any prior background in formal probability theory. We will introduce the basic concepts of probability here in a natural way as they emerge in the study of real problems.

7.1 Discrete Probability Models

The most simple and intuitive probability models are those involving a discrete set of possible outcomes and no time dynamic elements. Such models are frequently encountered in the real world.

Example 7.1. An electronics manufacturer produces a variety of diodes. Quality control engineers attempt to insure that faulty diodes will be detected in the factory before they are shipped. It is estimated that 0.3% of the diodes produced will be faulty. It is possible to test each diode individually. It is also possible to place a number of diodes in series and test the entire group. If this test fails, it means that one or more of the diodes in that group are faulty. The estimated testing cost is 5 cents for a single diode, and $4+n$ cents for a group of $n>1$ diodes. If a group test fails, then each diode in the group must be retested

Fourth edition

Variables:	$n =$ number of diodes per test group $C =$ testing cost for one group (cents) $A =$ average testing cost (cents/diode)
Assumptions:	If $n = 1$, then $A = 5$ cents Otherwise ($n > 1$), we have $C = 4 + n$ if the group test indicates that all diodes are good, and $C = (4 + n) + 5n$ if the group test indicates a failure. $A =$ (Average value of $C)/n$
Objective:	Find the value of n that minimizes A

Figure 7.1: Results of step 1 for the diode problem.

individually to find the bad one(s). Find the most cost–effective quality control procedure for detecting bad diodes.

We will use the five-step method. The results of step 1 are summarized in Figure 7.1. The variable n is a decision variable, and we are free to choose any $n = 1, 2, 3, \ldots$, but the variable C is the random outcome of the quality control procedure we select. We say C is a random variable. The quantity A, however, is not random. It represents the average or expected value of the random variable C/n.

Step 2 is to select the modeling approach. We will use a discrete probability model.

Consider a random variable X, which can take any of a discrete set of values

$$X \in \{x_1, x_2, x_3, \ldots\},$$

and suppose that $X = x_i$ occurs with probability p_i. We will write

$$\Pr\{X = x_i\} = p_i.$$

Of course, we must have

$$\Sigma p_i = 1.$$

Since X takes the value x_i with probability p_i, the average or expected value of X should be a weighted average of the possible values x_i, weighted according to their relative likelihoods p_i. We will write

$$EX = \Sigma x_i p_i. \tag{7.1}$$

The probabilities p_i represent what we will call the probability distribution of X.

Example 7.2. In a simple game of chance, two dice are rolled and the bank pays the player the number of dollars shown on the dice. How much would you pay to play this game?

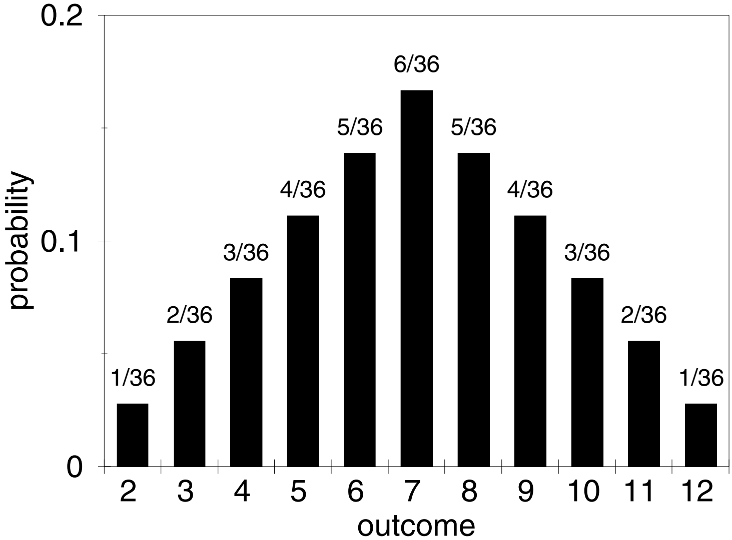

Figure 7.2: Histogram of probability versus outcome showing the distribution of the sum of two dice.

Let X denote the number shown on the dice. There are $6 \times 6 = 36$ possible outcomes, and each is equally likely. There is only one way to roll a 2, so

$$\Pr\{X = 2\} = 1/36.$$

There are two ways to roll a 3 (1 and 2, or 2 and 1), so

$$\Pr\{X = 3\} = 2/36.$$

The complete probability distribution of X is illustrated in Figure 7.2. The expected value of X is

$$EX = 2(1/36) + 3(2/36) + \cdots + 12(1/36),$$

or $EX = 7$. After many repetitions of this game, you would expect to win about seven dollars per roll. Therefore, it would be worthwhile to play the game if it cost no more than seven dollars to play.

To be more specific, suppose that you play the game over and over. Let X_n denote the amount you win on the nth roll of the dice. Each X_n has the same distribution, and the different X_n are independent. The amount won on one roll does not depend on the amount won in the previous roll. There is a theorem called "the strong law of large numbers," which says that for any sequence of independent, identically distributed random variables X, X_1, X_2, $X_3, \ldots$ with EX finite, we will have

$$\frac{X_1 + \cdots + X_n}{n} \to EX \tag{7.2}$$

as $n \to \infty$ with probability 1. In other words, if you play the game for a long time, you are virtually certain to win about \$7 per roll (Ross (1985) p. 70).

The formal definition of *independence* is as follows. Let Y and Z denote two random variables with

$$Y \in \{y_1, y_2, y_3, \ldots\}$$

and

$$Z \in \{z_1, z_2, z_3, \ldots\}.$$

We say that Y and Z are independent if it is generally true that

$$\Pr\{Y = y_i \text{ and } Z = z_j\} = \Pr\{Y = y_i\} \Pr\{Z = z_j\}. \tag{7.3}$$

For example, let Y and Z denote the number on the first and second dice, respectively. Then

$$\Pr\{Y = 2, Z = 1\} = \Pr\{Y = 2\} \Pr\{Z = 1\} = (1/6)(1/6) = (1/36),$$

and likewise for each possible outcome. The random variables Y and Z are independent. The number that comes up on the second die has nothing to do with what happened on the first.

Returning to the diode problem of Example 7.1, we see that the random variable C takes on one of two possible values for any fixed $n > 1$. If all the diodes are good, then

$$C = 4 + n.$$

Otherwise,

$$C = (4 + n) + 5n,$$

since we have to retest each diode. Letting p denote the probability that all the diodes are good, the remaining possibility (one or more bad diodes) must have probability $1 - p$. Then the average or expected value of C is

$$EC = (4 + n)p + [(4 + n) + 5n](1 - p). \tag{7.4}$$

Now for step 4. There are n diodes, and the probability that one individual diode is bad is 0.003. In other words, the probability that one individual diode is good is 0.997. Assuming independence, it follows that the probability that all n diodes in one test group are good is $p = 0.997^n$.

The expected value of the random variable C is

$$\begin{aligned} EC &= (4+n)0.997^n + [(4+n)+5n](1-0.997^n) \\ &= (4+n)+5n(1-0.997^n) \\ &= 4+6n-5n(0.997)^n. \end{aligned}$$

Hence, the average testing cost per diode is

$$A = \frac{4}{n} + 6 - 5(0.997)^n. \tag{7.5}$$

The strong law of large numbers tells us that this formula represents the long–run average cost we will experience if we use test groups of size n. Now all we need is to minimize A as a function of n. We leave the details to the reader (see Exercise 1). The minimum of $A = 1.48$ cents/diode occurs at $n = 17$.

We conclude with step 5. Quality control procedures for detecting faulty diodes can be made considerably more economical by group testing methods. Individual testing costs approximately 5 cents/unit. Bad diodes occur only rarely, at a rate of 3 per 1,000. By testing groups of 17 diodes each, in series, we can reduce testing costs by a factor of three (to 1.5 cents/diode) without sacrificing quality.

Sensitivity analysis is critical in this type of problem. The implementation of a quality control procedure will depend on several factors outside the scope of our model. It may be easier to test diodes in batches of 10 or 20, or perhaps n should be a multiple of 4 or 5, depending on the details of our manufacturing process. Fortunately, the average cost A does not vary significantly between $n = 10$ and $n = 35$. Again, we leave the details to the reader. The parameter $q = 0.003$, which represents the failure rate in the manufacturing process, must also be considered. For example, this value may vary with the environmental conditions inside the plant. Generalizing on our previous model, we have

$$A = \frac{4}{n} + 6 - 5(1-q)^n. \tag{7.6}$$

At $n = 17$ we have

$$S(A, q) = \frac{dA}{dq} \cdot \frac{q}{A} = 0.16,$$

so small variations in q are not likely to affect our cost very much.

A more general robustness analysis would consider the assumption of independence. We have assumed that there is no correlation between the times of successive failures in the manufacturing process. It may be, in fact, that bad diodes tend to be produced in batches, perhaps due to a passing anomaly in the manufacturing environment, such as a vibration or a power surge. The mathematical analysis of dependent random variable models cannot be dealt with in its entirety here. The stochastic process models introduced in the next chapter are capable of representing some kinds of dependence, while some other types of dependence admit no tractable analytic formulations. Problems in robustness

Variables:	λ = decay rate (per second) T_n = time of nth observed decay
Assumptions:	Radioactive decays occur at random with rate λ. $T_{n+1} - T_n \geq 3 \times 10^{-9}$ for all n
Objective:	Find λ on the basis of a finite number of observations $T_1, \ldots, T_n$

Figure 7.3: Results of step 1 of the radioactive decay problem.

are very much an active and intriguing branch of current research in probability theory. In practice, simulation results tend to indicate that expected value models based on independent random variables are quite robust. More importantly, it has been found through experience that such models provide useful, accurate approximations of real–life behavior in most cases.

7.2 Continuous Probability Models

In this section we consider probability models based on random variables that take values over a continuum. Such models are particularly convenient for representing random times. The mathematical theory required is completely analogous to the discrete case, except that now integrals replace sums.

Example 7.3. A "type I counter" is used to measure the radioactive decay in a sample of fissionable material. Decays occur at random, at an unknown rate, and the purpose of the counter is to measure the decay rate. Each radioactive decay locks the counter for a period of 3×10^{-9} seconds, during which time any decays that occur are not counted. How should the data received from the counter be adjusted to account for the lost information?

We will use the five-step method. The results of step 1 are summarized in Figure 7.3. Step 2 is to select the modeling approach. We will use a continuous probability model.

> Suppose that X is a random variable that takes values on the real line. A convenient way to describe the probability structure of X is to specify the function
>
> $$F(x) = \Pr\{X \leq x\},$$
>
> called the *distribution function* of X. If $F(x)$ is differentiable, we call the function
>
> $$f(x) = F'(x)$$
>
> the *density function* of X. Then for any real numbers a and b we have
>
> $$\Pr\{a < X \leq b\} = F(b) - F(a) = \int_a^b f(x)\,dx. \tag{7.7}$$

In other words, the area under the density curve represents probability. The mean or expected value of X is defined by

$$EX = \int_{-\infty}^{\infty} x f(x)\, dx, \tag{7.8}$$

which is directly analogous to the discrete case, as can be seen by considering the Riemann sum for the integral. (See Exercise 13 for details.) It is also worthwhile pointing out that this notation and terminology were originally adapted from a problem in physics, namely the center of mass problem. If a wire or rigid rod is laid out along the x axis, and $f(x)$ represents the density (gms/cm) at point x, then the integral of $f(x)$ represents mass, and the integral of $xf(x)$ represents the center of mass (assuming, as we do in probability, that the total mass is equal to 1).

The special case of random arrivals occurs frequently in applications. Suppose that arrivals (e.g., customers, phone calls, radioactive decays) occur at random with rate λ, and let X denote the random time between two successive arrivals. It is common to assume that X has the distribution function

$$F(t) = 1 - e^{-\lambda t}, \tag{7.9}$$

so that the density function of X is

$$f(t) = \lambda e^{-\lambda t}. \tag{7.10}$$

This distribution is called the *exponential distribution* with rate parameter λ.

One very important property of the exponential distribution is its "lack of memory". For any $t > 0$ and $s > 0$, we have

$$\begin{aligned} \Pr\{X > s + t | X > s\} &= \frac{\Pr\{X > s + t\}}{\Pr\{X > s\}} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} \\ &= \Pr\{X > t\}. \end{aligned} \tag{7.11}$$

In other words, the fact that we have already waited s units of time for the next arrival does not affect the (conditional) distribution of the time until the next arrival. The exponential distribution "forgets" that we have already waited this long. The probability in Eq. (7.11) is called a *conditional probability*. Formally, the probability of event A occurring, given the event B occurs, is

$$\Pr\{A|B\} = \frac{\Pr\{A \text{ and } B\}}{\Pr\{B\}}. \tag{7.12}$$

In other words, $\Pr\{A|B\}$ is the relative likelihood of A among all possible events once B has occurred.

Now we proceed to step 3, the model formulation. We are assuming that radioactive decays occur at random at an unknown rate λ. We will model this process by assuming that the times between successive radioactive decays are independent and identically distributed with an exponential distribution with rate parameter λ. Let

$$X_n = T_n - T_{n-1}$$

denote the times between successive observations of a radioactive decay. Of course, X_n does not have the same distribution as the time between successive decays, because of the lock time. In fact, $X_n \geq 3 \times 10^{-9}$ with probability 1, which is certainly not true for the exponential distribution.

The random time X_n consists of two parts. First, we must wait $a = 3 \times 10^{-9}$ seconds while the counter is locked, and then we must wait an additional Y_n seconds until the next decay. Now Y_n is not simply the time between two decays, because it begins at the end of the lock time, not at a decay time. However, the memoryless property of the exponential distribution guarantees that Y_n is still exponential with rate parameter λ.

Step 4 is to solve the model. Since $X_n = a + Y_n$, we have $EX_n = a + EY_n$, where

$$EY_n = \int_0^\infty t\lambda e^{-\lambda t}\, dt.$$

Integrate by parts to find $EY_n = 1/\lambda$. Thus, $EX_n = a + 1/\lambda$. The strong law of large numbers says that

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = a + \frac{1}{\lambda}$$

with probability 1. In other words, $(T_n/n) \to a + 1/\lambda$. For large n it will be approximately true that

$$\frac{T_n}{n} = a + \frac{1}{\lambda}. \tag{7.13}$$

Solving for λ, we obtain

$$\lambda = \frac{n}{T_n - na}. \tag{7.14}$$

Finally, we conclude with step 5. We have obtained a formula for decay rate that corrects for the decays missed while the counter is locked. All that is required is to record the length of observation and the number of decays recorded. The distribution of those decays in the observation interval is not required to determine λ.

Sensitivity analysis should consider the lock time a, which must be determined empirically. The accuracy to which we can determine a will affect the accuracy of λ. From Eq. (7.14) we calculate that

$$\frac{d\lambda}{da} = \lambda^2.$$

The sensitivity of λ to a is then

$$S(\lambda,\, a) = \lambda^2(a/\lambda) = \lambda a.$$

This is also the expected number of decays during the lock time. We can therefore get a better estimate of λ (in relative terms) for a less intensely radioactive source. One simple way to achieve this is to use fewer grams of radioactive material in our sample. Another important source of potential error comes from the assumption that

$$(X_1 + \cdots + X_n)/n = a + 1/\lambda.$$

Of course, this is not exactly true. Random fluctuations will cause the empirical rate to vary from the mean, although we do have convergence as $n \to \infty$. The study of such random fluctuations is the subject of the next section.

Finally, there is the matter of robustness. We have made an assumption about the decay process that appears to be very special. We have assumed that times between decays are independent and that they have a particular distribution (exponential with rate parameter λ). Such an arrival process is called a *Poisson process.* The Poisson process is commonly used to represent random arrivals. Its use can be justified in part by the fact that many real–world arrival processes have interarrival times that are at least approximately exponential. This can be verified by collecting data on arrival times. But this does not answer the question of *why* the exponential distribution occurs.

It turns out that there is a mathematical reason for expecting an arrival process to look Poisson. Consider a large number of arrival processes that are independent of one another. We make no assumption about the interarrival time distribution of an arrival process, only that the interarrival times are independent and identically distributed. There is a theorem that states, under fairly general conditions, that the arrival process obtained by merging all of these independent processes has to look Poisson. (The merged process tends to Poisson as the number of merged processes tends to infinity.) This is why the Poisson process, based on the exponential distribution, is such a robust model (see Feller (1971) p. 370).

7.3 Introduction to Statistics

In any modeling situation it is desirable to get quantitative measures of performance. For probability models an additional complication is involved in deriving such parameters of system behavior. We must have a way to deal with the random fluctuations in system behavior that are characteristic of probability models. Statistics is the study of measurement in the presence of random fluctuations. The appropriate use of statistical methods must be a part of the analysis of any probability model.

Example 7.4. An emergency 911 service in a local community received an average of 171 calls per month for house fires over the past year. On the basis of this data, the rate of house fire emergencies was estimated at 171 per month. The next month only 153 calls were received. Does this indicate an actual reduction in the rate of house fires, or is it simply a random fluctuation?

Variables:	λ = Rate of house fire reports (per month) X_n = time between $(n-1)$st and nth fire (months)
Assumptions:	House fires occur at random with rate λ; i.e., X_1, $X_2, \ldots$ are independent, and each X_n has an exponential distribution with rate parameter λ
Objective:	Determine the probability that as few as 153 calls would be received in one month, given $\lambda = 171$

Figure 7.4: Results of step 1 of the house fire problem.

We will use the five-step method. The results of step 1 are summarized in Figure 7.4. We are assuming exponential interarrival times for the emergency calls. Step 2 is to determine the modeling approach. We will model this as a statistical inference problem.

> Suppose that X, X_1, X_2, $X_3, \ldots$, are independent random variables, all with the same distribution. Recall that if X is discrete, the average or expected value is
>
> $$EX = \Sigma\, x_k \Pr\{X = x_k\},$$
>
> and if X is continuous with density $f(x)$, then
>
> $$EX = \int x f(x)\, dx.$$
>
> Another distributional parameter, called the *variance*, measures the extent to which X tends to deviate from the mean EX. In general we define
>
> $$VX = E(X - EX)^2. \tag{7.15}$$
>
> If X is discrete, we have
>
> $$VX = \Sigma(x_k - EX)^2 \Pr\{X = x_k\}, \tag{7.16}$$
>
> and if X is continuous with density $f(x)$, we have
>
> $$VX = \int (x - EX)^2 f(x)\, dx. \tag{7.17}$$
>
> There is a result called the *central limit theorem*, that states that as $n \to \infty$, the distribution of the sum $X_1 + \cdots + X_n$ gets closer and closer to a certain type of distribution called a *normal distribution.* Specifically, if we let $\mu = EX$ and $\sigma^2 = VX$, then for all t real we have
>
> $$\lim_{n\to\infty} \Pr\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le t\right\} \to \Phi(t) \tag{7.18}$$

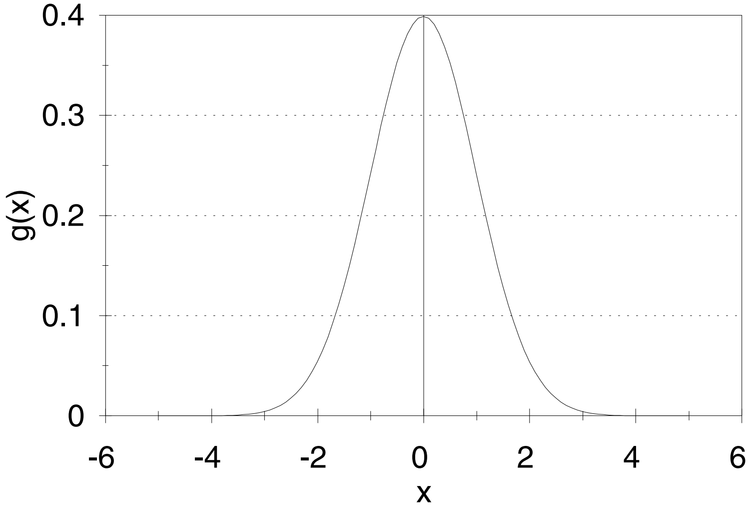

Figure 7.5: Graph of the standard normal density function (7.19).

where $\Phi(t)$ is a special distribution function called the *standard normal distribution*. The density function for the standard normal distribution is defined for all x by

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \tag{7.19}$$

so that, for all t, we have

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx. \tag{7.20}$$

Figure 7.5 shows a graph of the standard normal density. Numerical integration shows that the area between $-1 \le x \le 1$ is approximately 0.68, and the area between $-2 \le x \le 2$ is approximately 0.95. Thus, for all n sufficiently large, we will have

$$-1 \le \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le 1$$

about 68% of the time, and

$$-2 \le \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le 2 \tag{7.21}$$

about 95% of the time. In other words, we are 68% sure that

$$n\mu - \sigma\sqrt{n} \le X_1 + \cdots + X_n \le n\mu + \sigma\sqrt{n},$$

and 95% sure that

$$n\mu - 2\sigma\sqrt{n} \le X_1 + \cdots + X_n \le n\mu + 2\sigma\sqrt{n}. \qquad (7.22)$$

It is common in practice to accept the 95% interval from Eq. (7.22) as the range of normal variation in a random sample. In cases where the sum $X_1 + \cdots + X_n$ does not lie in the interval from Eq. (7.22), we say that the deviation is *statistically significant* at the 95% level.

Now we move on to step 3, the model formulation. We are assuming that the times between calls, X_n, are exponentially distributed, with density function

$$f(x) = \lambda e^{-\lambda x}$$

on $x \ge 0$. We have previously calculated that

$$\mu = EX_n = 1/\lambda.$$

The variance

$$\sigma^2 = VX_n$$

is given by

$$\sigma^2 = \int_0^\infty (x - 1/\lambda)^2 \lambda e^{-\lambda x}\, dx,$$

and the central limit theorem gives probability estimates of the extent to which

$$(X_1 + \cdots + X_n)$$

can vary from its mean, n/λ. In particular, we know that Eq. (7.22) holds with probability 0.95.

On to step 4. We calculate that

$$\sigma^2 = 1/\lambda^2,$$

using integration by parts. Substituting $\mu = 1/\lambda$ and $\sigma = 1/\lambda$ into Eq. (7.22), we find that the relation

$$\frac{n}{\lambda} - \frac{2\sqrt{n}}{\lambda} \le X_1 + \cdots + X_n \le \frac{n}{\lambda} + \frac{2\sqrt{n}}{\lambda} \qquad (7.23)$$

must hold with probability 0.95. Substituting $\lambda = 171$ and $n = 153$ into Eq. (7.23), we are 95% sure that

$$\frac{153}{171} - \frac{2\sqrt{153}}{171} \le X_1 + \cdots + X_{153} \le \frac{153}{171} + \frac{2\sqrt{153}}{171},$$

or in other words,

$$0.75 \le X_1 + \cdots + X_{153} \le 1.04.$$

Therefore, our observation that

$$X_1 + \cdots + X_{153} \approx 1$$

is within the range of normal variation.

Finally, step 5. There is insufficient evidence to conclude that the rate of house fire emergency calls has declined. The variation in the observed number of calls may well be the result of normal random variation. Of course, if the observed number of calls per month continues to be this low, then we would reassess the situation.

A few items should be included in our sensitivity analysis. First of all, we have concluded that 153 calls in one month is within the range of normal variation. More generally, suppose that n calls are received in one month. Substituting $\lambda = 171$ into Eq. (7.23), we conclude that

$$\frac{n}{171} - \frac{2\sqrt{n}}{171} \le X_1 + \cdots + X_n \le \frac{n}{171} + \frac{2\sqrt{n}}{171} \tag{7.24}$$

with probability 0.95. Since the interval

$$\frac{n}{171} \pm \frac{2\sqrt{n}}{171}$$

contains 1 for any value of $n \in [147, 199]$, we would conclude more generally that 95% of the time there will be between 147 and 198 calls in a month. In other words, the range of normal variation for this community is 147 to 198 emergency calls per month.

Now let us consider the sensitivity of our conclusions to the assumption that the actual expected number of emergency calls is 171 per month. Less specifically, assume that there is an average of λ emergency calls per month. We have observed $n = 153$ calls in a one–month period. Substituting into Eq. (7.23), we conclude that

$$\frac{153}{\lambda} - \frac{2\sqrt{153}}{\lambda} \le X_1 + \cdots + X_{153} \le \frac{153}{\lambda} + \frac{2\sqrt{153}}{\lambda} \tag{7.25}$$

with probability 0.95. Since the interval

$$\frac{153}{\lambda} \pm \frac{2\sqrt{153}}{\lambda}$$

contains 1 for any value of λ between 128 and 178, we conclude that a month with 153 emergency calls is within the range of normal variation for any community in which the average number of emergency calls is between 128 and 178 per month.

There is a final matter of robustness that requires comment. We have assumed that the times between calls, X_n, are exponential. However, the central limit theorem remains true for any distribution as long as μ and σ are finite. Hence, our conclusion is really not sensitive to the assumption of an exponential distribution. It only requires that σ is not too much smaller than μ (for an exponential distribution, $\mu = \sigma$). As we remarked at the end of Section 7.2, there is a good reason to expect this to be the case. Of course, we can always check by estimating μ and σ from the data.

7.4 Diffusion

The normal density function introduced in Section 7.3 is also important in another way. Brownian motion is the accumulation of small, independent random movements of particles. The sum of those independent random movements represents the location of the particle. The central limit theorem says that the distribution of this sum can be approximated by a normal probability density. Hence, the normal density is a model for the diffusion of small particles. In this section, we introduce the diffusion equation, a partial differential equation for particle spreading, in a way that highlights the close connection between the deterministic model and its probabilistic counterpart.

Example 7.5. An accident at an industrial plant ten kilometers upwind of a small town releases an airborne pollutant. One hour after release, a toxic cloud 2000 meters long is headed toward the town at a wind speed of 3 kilometers per hour. The maximum concentration of pollutant in the cloud is 20 times the safe level. What is the maximum concentration expected in town, when will it occur, and how long until the concentration of pollutant falls back below a safe level?

We will use the five-step method. The first step is to ask a question. We want to know the concentration of pollutant in town and how it varies over time. Let us assume that the cloud of pollutant moves at a constant velocity in the direction of the town. Because of diffusion, the pollutant will also spread as it moves, lowering the peak concentration. Hence, we expect the concentration to diminish with time. We will assume that the wind speed is constant at three kilometers per hour, so that the distance between the town and the plume center is decreasing at three kilometers per hour. We can use the fact that the cloud has spread to a length of 2000 meters in the first hour to estimate the rate of spreading. Then our objective is to predict how the concentration at the town location will vary over time as the cloud of pollutant passes through. The results of step 1 are summarized in Figure 7.6.

Step 2 is to select the modeling approach. We will use a diffusion model.

Diffusion is the spreading of particles due to small random movements. The relative concentration of contaminant $C(x,t)$ at location x at time t is described by a partial differential equation called the *diffusion equation*. Here the term *relative concentration* indicates that the concentration function has been normalized so that the total mass of particles $\int C(x,t)\,dx = 1$. This is useful to emphasize the connection to probability theory. The diffusion equation results from a combination of two elements. First of all, the law of conservation of mass states that

$$\frac{\partial C}{\partial t} = -\frac{\partial q}{\partial x} \tag{7.26}$$

where $q(x,t)$ is the *particle flux*, the number of particles passing through the point x per unit time. The change in concentration

Variables:	t = Time since release of pollutant (hrs) μ = Distance travelled by plume center (km) x = Distance between plume center and town (km) s = Plume spread at time t (km) P = Pollution concentration in town (times safe level)
Assumptions:	$\mu = 3t$ $x = 10 - \mu$ Peak concentration $P = 20$ when $t = 1$ hour Plume spread is $s = 2000$ meters at $t = 1$ hour
Objective:	Determine the maximum pollution level in town, and time until pollution falls back down to a safe level

Figure 7.6: Results of Step 1 of the pollution problem.

inside a small box of width Δx at location x is the result of the difference between the particle flux $q(x,t)$ into the box on the left side, and the flux $q(x+\Delta x, t) \approx q(x,t) + \Delta x\, \partial q/\partial x(x,t)$ out of the box at the right. If $\partial q/\partial x > 0$, then the mass is exiting on the right faster than it enters on the left, so the concentration will decrease. The net loss of mass over a time interval Δt is $\Delta M \approx -\Delta x\, \partial q/\partial x\ \Delta t$, and hence, the concentration $C = M/\Delta x$ changes by an amount $\Delta C = \Delta M/\Delta x \approx -\partial q/\partial x\, \Delta t$, which leads to $\Delta C/\Delta t \approx -\partial q/\partial x$. Taking limits as $\Delta x \to 0$ yields the conservation of mass equation (7.26). The second element is *Fick's Law*, the empirical observation that diffusive particle flux is proportional to the concentration gradient (particles tend to diffuse from areas of high concentration to areas of low concentration), or in other words,

$$q = -\frac{D}{2}\frac{\partial C}{\partial x} \tag{7.27}$$

for some constant $D > 0$, called the *diffusivity*. Combining (7.26) and (7.27) yields the diffusion equation

$$\frac{\partial C}{\partial t} = \frac{D}{2}\frac{\partial^2 C}{\partial x^2}, \tag{7.28}$$

which can be solved to predict the spread of contaminants.

The simplest way to solve the diffusion equation is to use *Fourier transforms*. The Fourier transform of a function $f(x)$ is given by the integral formula

$$\hat{f}(k) = \int_{-\infty}^{\infty} e^{-ikx} f(x) dx. \tag{7.29}$$

Using integration by parts, it is easy to show that the derivative $f'(x)$ has Fourier transform $(ik)\hat{f}(k)$. Tables of Fourier transforms are widely available, and computer algebra systems like Maple and Mathematica also compute Fourier transforms and their inverses (i.e., compute $f(x)$ for a given $\hat{f}(k)$ and vice versa). One useful formula that can be verified from tables or a computer algebra system is that the standard normal density (7.19) has Fourier transform

$$\hat{g}(k) = \int_{-\infty}^{\infty} e^{-ikx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = e^{-k^2/2}. \tag{7.30}$$

The standard normal distribution is connected with the diffusion model because sums of independent particle movements are asymptotically normal. Since we are interested in the way this distribution of particles spreads over time, we need to understand the scaling properties of the normal density. If Z is a standard normal random variable with mean zero and variance one, then $X = \sigma Z$ has mean zero and its variance is $E(X^2) = E(\sigma^2 Z^2) = \sigma^2$. The distribution function of X is given by

$$F(x) = P(X \le x) = P(\sigma Z \le x) = P(Z \le x/\sigma) = \int_{-\infty}^{x/\sigma} g(t)dt.$$

Substituting $u = \sigma t$ gives

$$F(x) = \int_{-\infty}^{x} g(\sigma^{-1}u)\sigma^{-1}du,$$

which shows that the random variable $X = \sigma Z$ has density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)}.$$

Then a change of variables $t = \sigma x$ in (7.30) shows that the Fourier transform of this density function is

$$\hat{f}(k) = e^{-(\sigma^2 k^2)/2}.$$

In the limiting case $\sigma \downarrow 0$ the Fourier transform is identically equal to one. This corresponds to a random variable concentrated at the origin, so that its spread is zero.

Now let us see how we can use Fourier transforms to solve the diffusion equation. Take Fourier transforms

$$\hat{C}(k,t) = \int_{-\infty}^{\infty} e^{-ikx} C(x,t)dx$$

in the diffusion equation (7.28) to obtain

$$\frac{d\hat{C}}{dt} = \frac{D}{2}(ik)^2\hat{C} = -\frac{D}{2}k^2\hat{C}, \tag{7.31}$$

which is a very simple ordinary differential equation ($u' = au$) for the Fourier transformed concentration. The solution to this differential equation with initial condition $\hat{C}(k,0) = 1$ for all k is $\hat{C}(k,t) = e^{-Dtk^2/2}$. This initial condition implies that at time $t = 0$ the contaminant plume is concentrated at the point $x = 0$ (the limiting case $\sigma \downarrow 0$). Now invert the Fourier transform to get

$$C(x,t) = \frac{1}{\sqrt{2\pi Dt}}\, e^{-x^2/(2Dt)}, \tag{7.32}$$

the point source solution to the diffusion equation (7.28).

The connection between the diffusion equation (7.28) and the central limit theorem from Section 7.3 can now be made explicit. Suppose that over a small time interval Δt a contaminant particle makes a small random movement X_i and that all of these movements are independent. Then the particle location at time $n\Delta t$ is given by the sum $X_1 + \cdots + X_n$ of these movements. Assume that the mean jump $E(X_i) = 0$ and the variance of a jump is $\sigma^2 = E(X_i^2)$. Then the central limit theorem (7.18) implies

$$\frac{X_1 + \cdots + X_n}{\sigma\sqrt{n}} \approx Z$$

standard normal, so that $X_1 + \cdots + X_n \approx \sigma\sqrt{n}Z$ in the sense of having approximately the same distribution. Hence, the probability distribution of particle location at time $t = n\Delta t$ is approximately Gaussian with variance $n\sigma^2 = (t/\Delta t)\sigma^2$. Taking $D = \sigma^2/\Delta t$, so that $n\sigma^2 = Dt$, we can relate the limiting random particle location $\sqrt{Dt}Z$ at time t with its probability density $C(x,t)$ in (7.32). For example, this relation shows that the contaminants spread at a rate proportional to $\sqrt{t}$. The normal density solution to the diffusion equation represents the relative concentration of contaminants, since it integrates to one. To represent concentration, it is only necessary to multiply by the total mass of contaminants.

Returning to the pollution problem, we continue with Step 3. The concentration P of pollutant is 20 times the safe level at time $t = 1$ hour and has spread to a width of $s = 2000$ meters. We assume a coordinate system where the pollutant release occurred at location 0 and the town is at location 10 kilometers. We also assume that the center of the plume has moved to location $\mu = 3t$ kilometers after t hours. Using the diffusion model, we can represent the contaminant plume at time $t = 1$ as a normal density with center or mean $\mu = 3$ kilometers and standard deviation $\sigma = 0.500$ kilometers, so that the interval $\mu \pm 2\sigma$ contains the bulk (about 95%) of the pollutant. In other words, we assume that $s = 4\sigma$. Then we can model the relative concentration of pollutant x kilometers away from the plume center of mass at time t by equation (7.32), where $D = \sigma^2 = 0.25$. This leads to the equation

$$C(x,t) = \frac{1}{\sqrt{0.5\pi t}}\; e^{-x^2/(0.5t)} \tag{7.33}$$

for the relative concentration x kilometers away from the plume center of mass at time t. Next, we note that the pollution concentration P is proportional to the relative concentration C. Using the units of % of safe level, we wish to set $P = 20$ when $t = 1$ and $x = 0$. Hence, we solve

$$P = 20 = P_0 \frac{1}{\sqrt{0.5\pi}}\; e^{-0^2/0.5}$$

to get $P = P_0 C$, where $P_0 = 20\sqrt{0.5\pi}$. Putting this all together, we get that the pollution level in the town, at time t and distance $x = 10 - \mu = 10 - 3t$ kilometers from the plume center of mass, is given by

$$P = \frac{20}{\sqrt{t}}\; e^{-(10-3t)^2/(0.5t)}, \tag{7.34}$$

and we want to answer the following questions: What is the maximum value of P and when does it occur? When will P fall below the safe level of $P = 1$?

Now we move on to Step 4. We want to maximize equation (7.34) over the set $t > 0$, and we want to solve the equation $P = 1$ for the largest positive root $t > 0$. We leave the details to the reader (see Exercise 15). The maximum of $P = 10.97$ occurs at the point $t_0 \approx 3.3$. There are two roots to the equation $P = 1$ over the set $t > 0$ that occur at (approximately) $t = 2.7$ and at $t = 4.1$.

Finally, for Step 5. The cloud of pollutant will expose the town to unsafe levels of contamination, almost 11 times the safe level. This peak level of risk will occur approximately 3 hours and 20 minutes after the pollution event. After around 4 hours the contamination level will fall back below safe levels. We also note that the level of contamination in the town will rise to an unsafe level for the first time at approximately 2 hours and 40 minutes after the accident.

Sensitivity analysis should focus on the parameters in our model that contain a significant amount of uncertainty. The biggest uncertainty is probably in the wind speed v, which could change over time. In our model we assumed $v = 3.0$ kilometers per hour. Generalizing the model, we now have that

$$P = \frac{20}{\sqrt{t}}\; e^{-(10-vt)^2/(0.5t)}, \tag{7.35}$$

and we repeat the analysis for different values of v. Table 7.1 shows the results of this exercise. It is clear that the wind speed has a very significant effect on the risk to the town. A higher wind speed produces a shorter exposure, but at a higher level of contamination. A lower wind speed does the opposite; it leads to a much longer exposure but at lower levels. Solving the problem again with a velocity of $v = 3.03$ (a 1% increase) yields a 0.5% increase in the maximum concentration M, so we conclude that $S(M, v) = 0.5$. The time T

Wind Speed (km/hr)	Highest Concentration	Time of Max Level (hrs)	Time Until Safe Level (hrs)
1.0	6.3	9.9	13.4
2.0	9.0	5.0	6.3
3.0	11.0	3.3	4.1
4.0	12.7	2.5	3.0
5.0	14.2	2.0	2.3

Table 7.1: Sensitivity analysis results for the pollution problem.

at which the maximum concentration M occurs goes down by -1.0 percent, so we conclude that $S(T, v) = -1$. The time L until the concentration falls to a safe level also decreases by 1% when the velocity increases by 1%, so we also have $S(L, v) = -1$. Of course, we could also compute these sensitivities from Table 7.1, with nearly the same results. We can also use the results of our sensitivity analysis to bound the parameters of risk in the event of variations in wind speed. For example, if the wind speed varies between 3 and 4 kilometers per hour over the next few hours, then we conservatively estimate that the maximum contamination in the town will be less than 12.7 times the safe level, and that contamination will drop to safe levels within 4.1 hours.

Finally, we come to the matter of robustness. Since the normal solution to the diffusion equation is connected to the central limit theorem, we expect a great deal of robustness. After all, independent random particle jumps have to converge to a normal distribution, regardless of the underlying distribution of the individual jumps. Another limitation of our model is that we assume the wind speed is the same at every point in space. It is true that the wind speed may vary, but we can easily account for this in our model by letting the velocity $v(t)$ vary over time. Then we can make the same kind of calculations if we are given accurate data or predictions of wind speed. In either case, the diffusion model predicts that a cloud of particles will spread away from its center of mass at a rate proportional to the square root of time. This is due to the fact that the particle density describes a normal random variable $\sqrt{Dt}Z$ with standard deviation (spread) $\sqrt{Dt}$. In many applications, it has been found that clouds of diffusing particles spread at a different rate than the classical model predicts. This is called *anomalous diffusion*, and it is a very active research area. See Exercise 18 for an illustration.

7.5 Exercises

1. Consider the diode problem of Example 7.1. Let

$$A(x) = \frac{4}{x} + 6 - 5(0.997)^x$$

denote the average cost function.

(a) Show that on the interval $x > 0$, $A(x)$ has a unique minimum value at the point where $A'(x) = 0$.

(b) Use a numerical method to estimate the minimum to within 0.1.

(c) Find the minimum value of $A(x)$ over the set $x = 1, 2, 3, \ldots$

(d) Find the maximum of $A(x)$ over the set $10 \leq x \leq 35$.

2. Reconsider the diode problem of Example 7.1. In this exercise we will investigate the problem of estimating the failure rate q. In Example 7.1 we assumed that $q = 0.003$; in other words, that 3 out of 1,000 diodes are flawed.

   (a) Suppose that we test a batch of 1,000 diodes and we find that 3 are flawed. On this basis we estimate that $q = 3/1,000$. How accurate is this estimate? Use the five-step method, and model as a statistical inference problem. [Suggestion: Define a random variable X_n to represent the status of the nth diode. Let $X_n = 0$ if the nth diode is good, or $X_n = 1$ if the nth diode has a flaw. Then the sum

$$\frac{X_1 + \cdots + X_n}{n}$$

   represents the fraction of bad diodes, and the central limit theorem can be used to estimate the likely variation of this sum from the true mean q.]

   (b) Repeat part (a), assuming that 30 bad diodes were found in a batch of 10,000.

   (c) How many diodes need to be tested in order to be 95% sure that we have determined the failure rate to within 10% of its true value? Assume that the true value is close to our original estimate of $q = 0.003$.

   (d) How many diodes need to be tested in order to be 95% sure that we have determined the failure rate to within 1% of its true value? Assume that the true value is close to our original estimate of $q = 0.003$.

3. Consider the radioactive decay problem of Example 7.3, and suppose that our counter registers an average of 10^7 decays/sec over a period of 30 seconds.

   (a) Use Eq. (7.14) to estimate the actual decay rate λ.

   (b) Generalize Eq. (7.13), using the central limit theorem. Calculate the range of normal variation (at the 95% level) for the observed decay rate T_n/n for an arbitrary value of the true decay rate λ.

   (c) Determine the range of λ for which the observed decay rate $T_n/n = 10^7$ is within the range of normal variation at the 95% level. How accurate is our estimate in part (a)?

(d) How long would we have to sample this radioactive material to be 95% sure we have determined the true decay rate λ to six significant digits (i.e., to within $0.5 \times 10^{-6}\,\lambda$)?

4. It can be shown that the area beneath the graph of the standard normal density function (see Eq. (7.19)) between $-3 \le x \le 3$ is approximately 0.997. In other words, for large n we are about 99.7% sure that

$$n\mu - 3\sigma\sqrt{n} \le X_1 + \cdots + X_n \le n\mu + 3\sigma\sqrt{n}.$$

Use this fact to repeat the calculations in Exercise 3, but use the 99.7% confidence level. Comment on the sensitivity of your answer in part (d) to the confidence level.

5. Reconsider the house fire problem of Example 7.4. In this exercise we will investigate the problem of estimating the rate λ at which emergency calls occur.

   (a) Suppose that 2,050 emergency calls are received in a one–year period. Estimate the rate λ of house fires per month.

   (b) Assuming that the true value of λ is 171 calls per month, calculate the range of normal variation for the number of emergency calls received in one year.

   (c) Calculate the range of λ for which 2,050 calls in one year is within the range of normal variation. How accurate is our estimate of the true rate λ at which house fires occur?

   (d) How many years of data would be required to obtain an estimate of λ accurate to the nearest integer (an error of ± 0.5)?

6. Reconsider the house fire problem of Example 7.4. The underlying random process is called a Poisson process because it can be shown that the number of arrivals (calls) N_t during a time interval of length t has a Poisson distribution. Specifically,

$$\Pr\{N_t = n\} = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

for all $n = 0,\, 1,\, 2, \ldots$

   (a) Show that

$$EN_t = \lambda t$$

   and

$$VN_t = \lambda t.$$

   (b) Use the Poisson distribution to calculate the probability that the number of calls received in a given month deviates from the mean of 171 by as much as 18 calls.

(c) Generalize the calculation of part (b) to determine the exact range of normal variation (at the 95% level) for the number of calls in a one–month period.

(d) Compare the exact method used in part (c) with the approximate calculation of the range of normal variation that is included in the discussion of sensitivity analysis for Example 7.4 in the text. Which method would be more appropriate for determining the range of normal variation in the number of calls received in a single day? In a year?

7. The Michigan state lottery runs a game in which you pay \$1 to buy a ticket containing a three–digit number of your choice. If your number is drawn at the end of the day, you win \$500.

(a) Suppose you were to buy one ticket per week for a year. What are your chances of coming out a winner for the year? [Hint: It is easy to compute the probability of coming out a loser!]

(b) Can you improve your chances of coming out a winner this year by purchasing more than one ticket per week? Calculate the probability of coming out a winner if you buy n tickets a week, for $n = 1, 2, 3, \ldots, 9$.

(c) Suppose that the state lottery sells 1,000,000 tickets this week. What is the range of likely variation in the amount of money the state will make this week? How likely is it that the state will lose money this week? Use the central limit theorem.

(d) What is wrong with using the central limit theorem to answer the question in part (a) or (b)?

8. (Murphy's Law, part I) You are staying at a downtown hotel. In front of the hotel there is a taxicab stand. Taxis arrive at random, at a rate of about one every five minutes.

(a) How long do you expect to wait for a taxicab, assuming that there is none at the hotel when you exit?

(b) The time until the arrival of the next taxicab is called a *forward recurrence time*. The time since the most recent arrival is called a *backward recurrence time*. For a Poisson process it can be shown that backward and forward recurrence times have the same distribution. (The probabilistic behavior of the process is the same if we let time run in reverse.) Using this fact, how long on average has it been since the last taxi arrived at the time you exit the hotel?

(c) On average, the time between taxicab arrivals is five minutes. How long on average is the length of time between the arrival of the cab you just missed and the one you have to wait for?

9. (Murphy's Law, part II) You are at the supermarket checkout stand. After waiting for what seems like an unusually long time to check out, you decide to conduct a scientific experiment. One by one, you measure the length of time each customer has to wait. You continue until you find one who has to wait longer than you did.

   (a) Let X denote the time you had to wait, and let X_n denote the time the nth customer had to wait. Let N denote the number of the first customer n for which $X_n \geq X$. To be fair, assume that X, X_1, $X_2, \ldots$ are identically distributed. Explain why the probability that $N \geq n$ (i.e., that out of the group consisting of you and the first $n-1$ customers, you waited the longest) must equal $1/n$.

   (b) Calculate the probability distribution of the random variable N.

   (c) Calculate the expected value EN, which represents the number of customers you must observe, on average, until you find one who waited longer than you did.

10. (Murphy's Law, part III) A doctor of internal medicine with a busy practice expects to be called into the hospital to respond to a serious heart attack on average about once every two weeks. Assume that heart attacks in this physician's patient population occur at random with this rate. One such emergency call is a challenge. Two such calls in a single day is a disaster.

    (a) How many heart attacks should the physician expect to respond to in a single year?

    (b) Explain why the probability that n heart attacks during the year all occur on different days is

$$\frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365-n+1}{365}.$$

    (c) What is the probability that the doctor has to respond to two or more heart attacks on the same day sometime this year?

11. A squadron of 16 bombers needs to penetrate air defenses to reach its target. They can either fly low and expose themselves to the air defense guns, or fly high and expose themselves to surface–to–air missiles. In either case, the air defense firing sequence proceeds in three stages. First, they must detect the target, then they must acquire the target (lock on target), and finally they must hit the target. Each of these stages may or may not succeed. The probabilities are as follows:

AD Type	P_{detect}	P_{acquire}	P_{hit}
Low	0.90	0.80	0.05
High	0.75	0.95	0.70

The guns can fire 20 shells per minute, and the missile installation can fire three per minute. The proposed flight path will expose the planes for one minute if they fly low, and five minutes if they fly high.

(a) Determine the optimal flight path (low or high). The objective is to maximize the number of bombers that survive to strike the target.

(b) Each individual bomber has a 70% chance to destroy the target. Use the results of part (a) to determine the chances of success (target destroyed) for this mission.

(c) Determine the minimum number of bombers necessary to guarantee a 95% chance of mission success.

(d) Perform a sensitivity analysis with respect to the probability $p = 0.7$ that an individual bomber can destroy the target. Consider the number of bombers that must be sent to guarantee a 95% chance of mission success.

(e) Bad weather reduces both P_{detect} and p, the probability that a bomber can destroy the target. If all of these probabilities are reduced in the same proportion, which side gains an advantage in bad weather?

12. A scanning radio communications sensor attempts to detect radio emissions and pinpoint their locations. The sensor scans 4,096 frequency bands. It takes the sensor 0.1 seconds to detect a signal. If no signal is detected, it moves to the next frequency. If a signal is detected, it takes an additional five seconds to get a location fix. There is no signal except on about 100 of the frequency bands, but the sensor does not know which ones are being used, so it must scan them all. On the busy frequencies, the percent utilization (i.e., the fraction of time that the signal is on) varies from 30% to 70%. An additional complication is that emissions on the same frequency can come from several different sources, so the sensor must continue to scan all frequencies even after a source is located.

(a) Determine the approximate detection rate for this system. Assume that all frequency bands are scanned sequentially.

(b) Suppose the sensor has the capability of remembering 25 high–priority frequency bands, which are scanned ten times as often as the others. Assuming that the sensor is eventually able to identify 25 busy frequencies and gives them high priority, how does the detection rate change?

(c) Suppose that, in order to obtain useful information, we must be able to detect emissions on a particular frequency at a minimum rate of once every three minutes. Determine the optimal number of high–priority channels.

(d) Perform a sensitivity analysis on the answer to part (c) with respect to the average utilization rate for busy frequencies.

13. In this problem we will explore the parallel between discrete and continuous random variables. Suppose that X is a continuous random variable with distribution function $F(x)$ and density function $f(x) = F'(x)$. For each n we define a discrete random variable X_n with approximately the same distribution as X. Partition the real line into intervals of length $\Delta x = n^{-1}$, and let I_i denote the ith interval in the partition. Then, for each i, select a point x_i in the ith interval, and define

$$p_i = \Pr\{X_n = x_i\} = f(x_i)\Delta x.$$

   (a) Explain why we can always choose the points x_i so that we will have

   $$p_i = \Pr\{X \in I_i\}$$

   for all i. This ensures that $\sum p_i = 1$.

   (b) Derive a formula that represents the probability that $a < X_n \leq b$ in terms of the density function f for any two real numbers a and b.

   (c) Derive a formula that represents the mean EX_n in terms of the density function f.

   (d) Use the results of part (b) to show that as $n \to \infty$ (or, equivalently, $\Delta x \to 0$), we have

   $$\Pr\{a < X_n \leq b\} \to \Pr\{a < X \leq b\}$$

   for any two real numbers a and b. We say that X_n *converges in distribution* to X.

   (e) Use the results of part (c) to show that as $n \to \infty$ (or, equivalently, $\Delta x \to 0$), we have

   $$EX_n \to EX.$$

   We say that X_n *converges in mean* to X.

14. (Geometric distribution) In this problem we investigate the discrete version of the exponential distribution. Suppose that arrivals occur at random at times $i = 1, 2, 3, \ldots$ Then the time X between two successive arrivals has the geometric distribution

$$\Pr\{X = i\} = p(1-p)^{i-1},$$

where p is the probability of an arrival occurring at time i.

   (a) Show that $\Pr\{X > i\} = (1-p)^i$. [Hint: Use the geometric series $1 + x + x^2 + x^3 + \cdots = (1-x)^{-1}$.]

   (b) Use part (a) to show that X has the memoryless property $\Pr\{X > i + j | X > j\} = \Pr\{X > i\}$.

   (c) Compute $EX = 1/p$. [Hint: Differentiate the geometric series to obtain $1 + 2x + 3x^2 + \cdots = (1-x)^{-2}$.] Explain why p is the arrival rate for the discrete process.

(d) Customers arrive at a public telephone at random at the rate of one every ten minutes. If Y is the time of first afternoon arrival, use the exponential model to compute $\Pr\{Y > 5\}$. If X is the number of minutes until the next arrival (take the clock time Y and just record the minutes X), use the geometric model to compute $\Pr\{X > 5\}$. Compare.

15. Consider the pollution problem of Example 7.5. Let

$$P(t) = \frac{20}{\sqrt{t}}\, e^{-(10-3t)^2/(0.5t)}$$

denote the pollution concentration in town at time t.

(a) Graph the function $P(t)$ and comment on its important features.

(b) Show that on the interval $t > 0$, $P(t)$ has a unique maximum at the point where $P'(t) = 0$.

(c) Use a numerical method to estimate the maximum to within 0.1.

(d) Show that the equation $P(t) = 1$ has two positive roots.

(e) Use a numerical method to estimate each positive root of the equation $P(t) = 1$ to within 0.1.

16. A chemical spill has contaminated the ground water near a municipal well. One year after the spill, the contaminant plume is 500 meters downstream of the spill, with a width of 200 meters. The concentration at the center of the plume is 3.6 parts per million.

(a) How long will it take until the maximum concentration reaches the municipal well located 1,800 meters downstream? What will the concentration be? Assume a one–dimensional diffusion model with constant velocity, as in Example 7.5.

(b) When will the concentration at the municipal well fall below a safe level of 0.001 parts per million?

(c) Compute the sensitivity of the answers in parts (a) and (b) to the groundwater velocity.

(d) Compute the sensitivity of the answers in parts (a) and (b) to the measured width of the contaminant plume.

17. (2–D pollution problem) In this problem we consider a two–dimensional diffusion model for the pollution problem of Example 7.5. The relative concentration $C(x, y, t)$ at location (x, y) at time t follows a bivariate normal distribution

$$C(x, y, t) = \frac{1}{\sqrt{2\pi Dt}}\, e^{-x^2/(2Dt)} \cdot \frac{1}{\sqrt{2\pi Dt}}\, e^{-y^2/(2Dt)},$$

where the center of the plume is assumed to be at location $x = 0$, $y = 0$. This bivariate density function solves the 2-D diffusion equation

$$\frac{\partial C}{\partial t} = \frac{D}{2}\frac{\partial^2 C}{\partial x^2} + \frac{D}{2}\frac{\partial^2 C}{\partial y^2}$$

and the connection to the probability model is the same as before, except that we also consider the spreading effect of small random jumps in the y direction. As in Example 7.5, an accident at an industrial plant ten kilometers upwind of a small town releases an airborne pollutant. One hour after release, a toxic cloud 2,000 meters wide is headed toward the town at a wind speed of three kilometers per hour, and the peak concentration is 20 times the safe level. Assume that the wind is blowing in the positive x direction, so that the plume center of mass is at location $(3, 0)$ at time $t = 1$, and the town is at location $(10, 0)$.

(a) Make a 3–D plot of the concentration plume at time $t = 1$. Comment on its most important features.

(b) How long will it take until the maximum concentration reaches the town located ten kilometers downwind? What will the concentration be?

(c) When will the concentration in the town fall below a safe level of 0.001 parts per million?

(d) Repeat parts (b) and (c), assuming that the wind is not blowing directly into town. Assume that at time $t = 1$ hour the plume center of mass is at location $(2.95, 0.5)$. How much difference does the wind direction make?

(e) Compare the results of parts (b) and (c) to what was found in the text. Does it make a significant difference whether we use a 1-D or a 2-D diffusion model?

18. (Anomalous diffusion) In this exercise we will investigate a model for anomalous super-diffusion, where a plume spreads faster than the classical diffusion equation predicts. Reconsider the pollution problem of Example 7.5, but now assume that the dispersivity $D(t)$ grows with time, so that $D(t) = 0.25t^{0.4}$ in Eq. (7.32).

(a) Repeat the calculations of Example 7.5. How long will it take until the maximum concentration reaches the town? What will the concentration be?

(b) When will the concentration in town fall to a safe level?

(c) Examine the sensitivity of the answers in parts (a) and (b) to the scaling parameter $p = 0.4$. Repeat parts (a) and (b) for $p = 0.2$, 0.3, 0.5, and 0.6, and discuss.

(d) Compare the results of parts (b) and (c) to what was found in the text. How does the possibility of anomalous super-diffusion affect the conclusions in the text?

Further Reading

1. Barnier, W. *Expected Loss in Keno.* UMAP module 574.
2. Berresford, G. *Random Walks and Fluctuations.* UMAP module 538.
3. Billingsley, P. (1979) *Probability and Measure.* Wiley, New York.
4. Carlson, R. *Conditional Probability and Ambiguous Information.* UMAP module 391.
5. Feller, W. (1971) *An Introduction to Probability Theory and Its Applications.* Vol. 2, 2nd ed., Wiley, New York.
6. Moore, P. and McCabe, G. (1989) *Introduction to the Practice of Statistics.* W.H. Freeman, New York.
7. Ross, S. (1985) *Introduction to Probability Models.* 3rd ed., Academic Press, New York.
8. Watkins, S. *Expected Value at Jai Alai and Pari-Mutuel Gambling.* UMAP module 631.
9. Wheatcraft, S. and Tyler, S. (1988) An explanation of scale-dependent dispersivity in heterogeneous aquifers using concepts of fractal geometry. *Water Resources Research* **24**, 566–578.
10. Wilde, C. *The Poisson Random Process.* UMAP module 340.